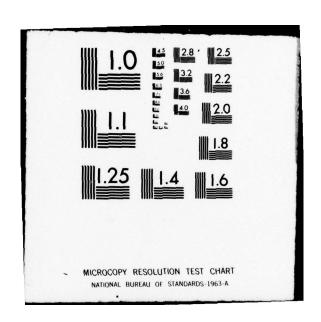
TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS
TIME SERIES ANALYSIS OF LH HORMONE LEVELS. (U)
FEB 79 H J NEWTON
N00014-AD-A067 771 F/6 6/1 N00014-78-C-0599 UNCLASSIFIED NL AD A06 771 器 RH END DATE FILMED 6 -79 DDC



TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS 77843

SITT IF OF STATISTICS New 713 - 545-3141

177730A0A

Technical Report No. N-5

February 1979

Institute of Statistics, Texas A&M University

By H. Joseph Newton

Texas A & M Research Foundation Project No. 3838

Sponsored by the Office of Naval Research "Multiple Time Series Modeling and Time Series Theoretic Statistical Mathods"

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

FILE COPY

300

Technica] Time Series Analysis of LH Hormone Levels, ATION PAGE TITLE (and But

PROGRAM ELEMENT, PROJECT, TASK CONTRACT OR GRANT NUMBERY ND6614-78-C-6599

CONTROLLING OFFICE MANE AND ADD

Arlington, VA 22217

Texas A&M University

H. Joseph

TIME SERIES ANALYSIS OF LH HORMONE LEVELS

S. SECURITY CLASS. fol this re Unclassified

A. DECLASSIFICATION DOWNGRADING

ARR 20 1979

Approved for public release; distribution unlimited.

×

SUPPLEMENTARY HOTE!

D. KEY WORDS (Continue on reverse olds if necessary and identify by block Empirical time series analysis

to data from the Physiology of Animal Reproduction. The presentation is intended to be nontechnical in nature so that researchers in a wide variety of fields can infer the applicability of the methods. This report illustrates the methods of time series analysis as applied

The output of a computer program written by the author is described.

DD 1 JAN 73 1473 EDITION OF 1 HOV 65 IS DESOLETE S/N 0102- LF-014-6601

Unclassified

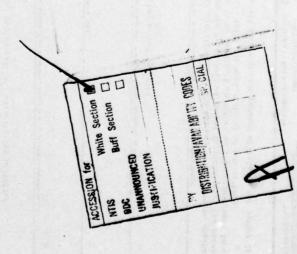
347380

THE SERIES ANALYSIS OF LH HORMONE LEVELS R. Joseph Newton Institute of Statistics, Texas AAM University

.

Abstract

This report illustrates the methods of time series analysis as applied to data from the Physiology of Animal Reproduction. The presentation is intended to be nontechnical in nature so that researchers in a wide variety of fields can infer the applicability of the methods. The output of a computer program written by the author is described.



1. Introduction

This report describes a set of data consisting of levels of the hormone LH measured at ten minute intervals for 24 hours in four covs each on three different days. Thus there are twelve time series each having 144 observations. Included in the appendix are plots of the data sets. Plots are included on an overall scale (to compare level of the series) and on an individual scale (to compare types of variability in the series).

Section 2 describes some questions of interest concerning the time series and outlines the statistical methodology used. In Section 3 initial results are presented.

2. Aims of the Analysis

These time series present many interesting questions. In this report the following particular aspects are investigated:

- What is the periodic behavior of the various series; i.e., can the variability in a series be explained by sinusoidal behavior of various periods?
- 11) Are mean levels of LN significantly different for the different series?
- 111) Can one of the standard time series models be used to fit the data and predict future values of the series?

To answer these questions a brief description of the basic methodology of time series analysis is necessary. What distinguishes time series data X(1), ..., X(T) from the type of data traditionally considered in statistics is that individual observations are not independent of the others. We

that as v increases, R(v) approaches zero, i.e., observations far apart that as v increases, R(v) approaches zero, i.e., observations far apart in time are almost independent. Thus one description of a time series is the <u>correlogram</u> p(v) vs. v where p(v) = R(v)/R(0) measures the correlation of data points v spart.

Another useful tool is to consider the discrete Pourier transform of the sequence of R(v). This is the function f(v), v c [0, 1] which is called the spectral density function of the time series. This function is used roughly in the same way an electrical engineer uses harmonic analysis. Thus the plot of a time series is similar to a waveform and the spectral density function assesses the amount of variability in the series due to cosines and sines in the various frequency ranges.

Time series methods based on the correlogram are called <u>time domain</u>
<u>sethods</u> while methods based on the spectral density are called <u>frequency</u>
<u>domain methods</u>. Many analysts use only time domain methods while we prefer
to use the insights gained by using both types.

The computer program used in analyzing a time series essentially consists of estimating the correlogram by its sample analog and finding three different estimators of the spectral density; (1) periodogram $f_{T,M}(v)$, and (111) autoregressive estimator $f_T(v)$, (11) smoothed periodogram $f_{T,M}(v)$, and (111) autoregressive estimator $f_T(v)$. The autoregressive estimator also gives a linear statistical model for describing the data. An observation X(v) is assumed to be a linear combination of the previous p observations in time plus a random error. The program estimates a p which is best in some sense and the weights $\tilde{a}(1)$, ..., $\tilde{a}(\tilde{p})$ of the linear combinations.

÷

3. Results

Computer Output

For each data set there are twelve plots. The series is identified by the caption above the plot while the plot is identified (except for the last two) by the label below the horizontal axis. The plots are described as follows:

Identifier	Plot
DATA	Data set
E	Periodogram estimator
CORR	Correlogram estimator
E	Smoothed estimator f _{T,M} using the Parzen wind
	for 3 orders M of smoothing
PART	Partial autocorrelations
BRV	Blased Residual Variance Estimator
URV	Unblased Residual Variance Estimator
CAT	Parzen's CAT criterion
ARSB	Autoregressive Estimator, Best Order
ARS2	Autoregressive Estimator, 2nd Best Order

The last two plots show the data (solid line) and values fitted by the autoregressive model (x's). The plots PART, BRV, URV, and CAT help identify the autoregressive order $\hat{p}.$

Periodic Behavior

The CORR plot is used to suggest important periodicities although it is important to realize that such periodicities can be misleading. If

the correlogram is high at a certain value, it could be that a periodicity of that value exists.

Peaks in the FT, FH, ANSB, and ANS2 plots at a frequency suggest a periodicity of period the reciprocal of the frequency. The ANSB and ANS2 plots actually have the periods for the peaks written out.

Thus for series Animal 1, day 1, (identified as HCLA 11) the correlogram is high at lag 7 while the peaks in the spectra occur at frequencies around 1/7.

If one suspects that a periodicity of 1/v exists, this suspecton can be tested by rejecting the hypothesis of no periodicity at the $\alpha=.05$ level if $\frac{f_{T,H}(v)}{f_{T}(v)}<\frac{1}{3}\;(1+\frac{3}{2}\frac{H}{T})=\frac{5}{12}$. Here M is the order of smoothing and T is the length of the series; we used M = $\frac{1}{6}$. Thus for series (1, 1) at frequencies in the region close to 1/7 the hypothesis of no periodicity is clearly rejected.

The results of the investigation into periodic behavior are summarized:

Periodic Structure of LH Level Series

		y signifi-	y signifi-			k around 6.		
SPEC	Significant peak around 7	High low frequency and possibly signifi- cant peak in 20-25 range	High low frequency and possibly significant peak in 20-25 range	Significant peak around 4	Similar to 1, 2	All behavior dominated by low frequency. Maybe peak around 6.	Significant peak around 5	
CORR		Decays quickly	Decays quickly	•	Similar to 1, 2	All behavior dominat	•	
les		2		1	2	3		
Series	1,1	-	1,3	2, 1	7,	2,	3,	

-5-

	00.7
	cont.
	Ser les
	LA Level
	10
	otructure c
Davidde	TENT TOT

	ier toute office of	retront structure of th Level Series (cont.)	
Series	CORR	SPEC	
3, 3	Same as 2, 3 except	Same as 2, 3 except maybe peak around 8 or 9.	
	•	Significant peak around 5. High frequency peak might be significant.	
4, 2	Similar to 1, 2	Similar to 1, 2	
£.3	**	Maybe significant peak around 14.	

ean Level of Serie

To make inferences about a population mean μ from a sample mean \bar{x} , one needs $\sigma_{\bar{x}}^2$, the variance of \bar{x} . For independent, identically distributed (iid) random variables (i.e., p=0), $\sigma_{\bar{x}}^2=\frac{\sigma^2}{T}$ where σ^2 is the variance of the data. Correlated data is modeled as an autoregression of order p giving approximately $\sigma_{\bar{y}}^2=\frac{\sigma^2}{T}C$, where

and σ^2 is the variance of the random error in the model. Thus for large samples we conclude that the means of series j and series k are significantly different if

where $D_{jk}=2(\sqrt{\hat{\sigma}_{k}^{2}}+\hat{\sigma}_{k}^{2})$ where $\hat{\sigma}_{k}^{2}$ is the same as $\hat{\sigma}_{k}^{2}$ with estimators \hat{p} , $\hat{\alpha}(1)$, ..., $\alpha(\hat{p})$, $\hat{\sigma}^{2}$ replacing the true unknown p, $\alpha(1)$, ..., $\alpha(p)$, σ^{2} .

We summarize these below for the first and second series for each animal. The third series are not included because the approximations to the $\sigma_{\rm X}^2$ are not valid.

.

Series	111	92		93
	;	:	1:	1
1.1	17.16	6.30	13	.489
	15.50	24.75	7	1.189
	12.72	4.31	18	1.061
	11.05	18.46	2	1.096
	15.35	5.65	5	.196
	12.28	29.74	1	1.695
	22.67	11.53	1	.126
	22.20	144.76	7	9.584
Differences	ences and Critical Values for Comparing Means	Values for	Comparing	Means
1.1 1.2	2 2.1	2. 2 3. 1		3. 2

10.39 7.32 2.75 1.23 11.62 11.15 2.63 3.32 9.95 3.02 2.35 3.40 7.17 -6.11 2.96 5.51 2.49 -1.81 4, 2 3, 1

4

It is important to realize that this table is similar to a multiple comparisons procedure in analysis of variance; A.e., the probability of all of these tests leading to correct decisions is very small.

Models

Prom inspection of the final plot for each series it appears that the autoregressive models fit the data well and provide evidence discriminating between series.

Conclueion

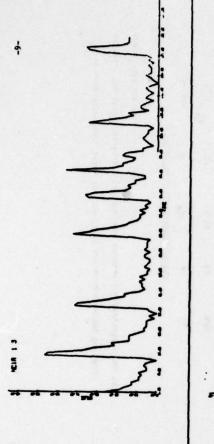
It is obvious that further investigation is necessary to lead to more precise conclusions. However, it is clear from this initial work that these data exhibit some periodic behavior. Further, conclusions can be reached about mean levels of the series and that the autoregressive models can assist in understanding the statistical and physical mechanisms generating the observations.

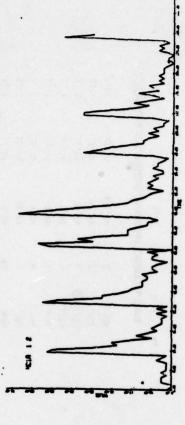
know ledgement

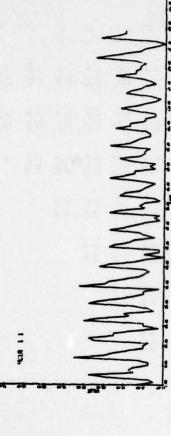
The author vishes to thank Professor P. Harms of the Department of Amismi Science at Texas A&M University for providing the data analyzed in this report.

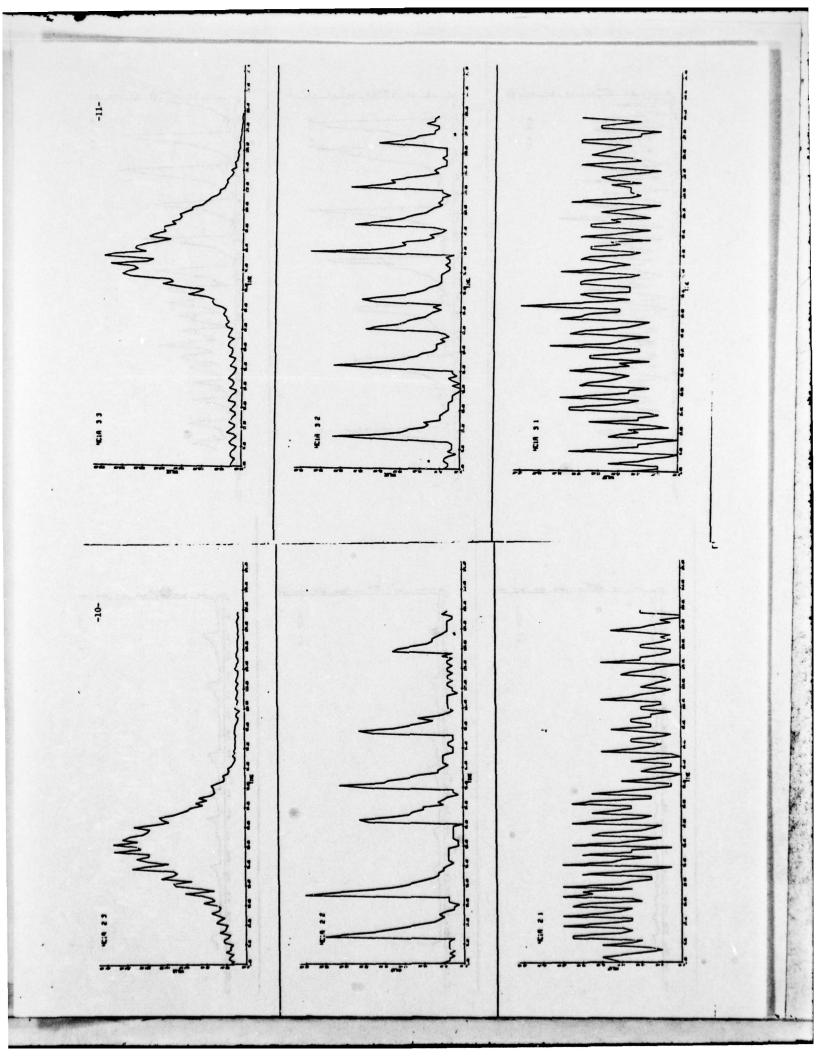
Reference:

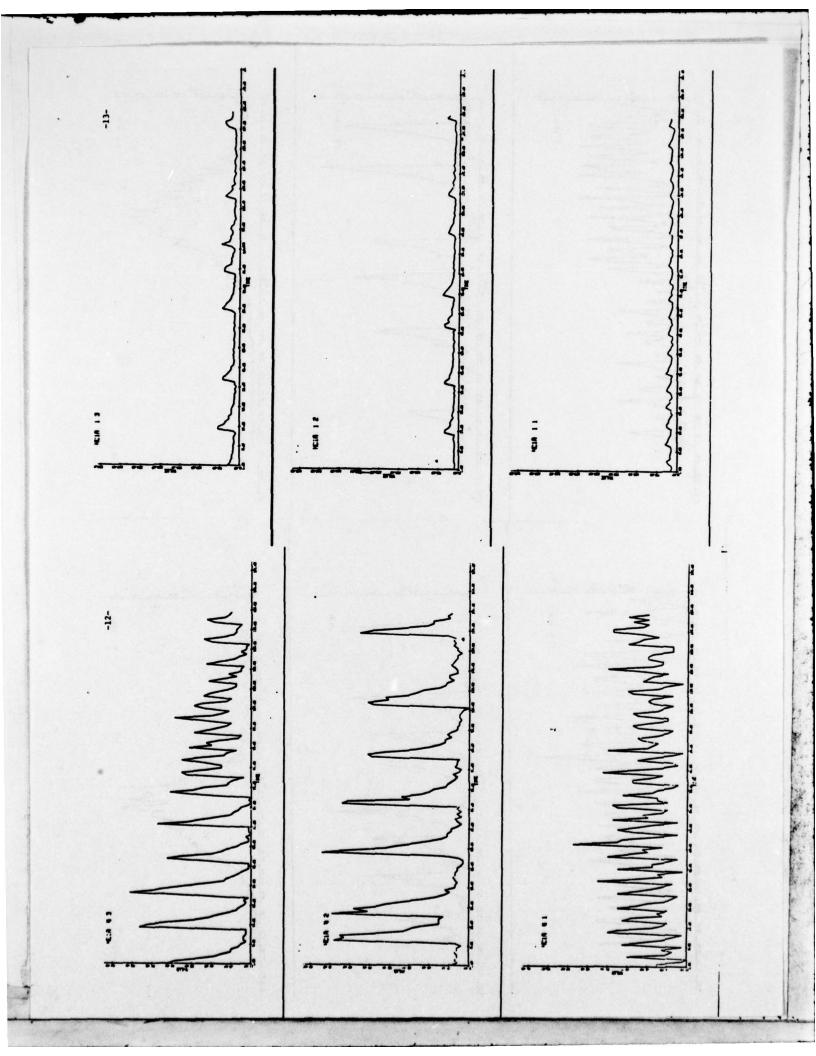
For a more technical statistical description of the methods in this report, see "Time Series Modelling, Spectral Analysis, and Forecasting" by E. Parzen, Report N-7, March 1979.

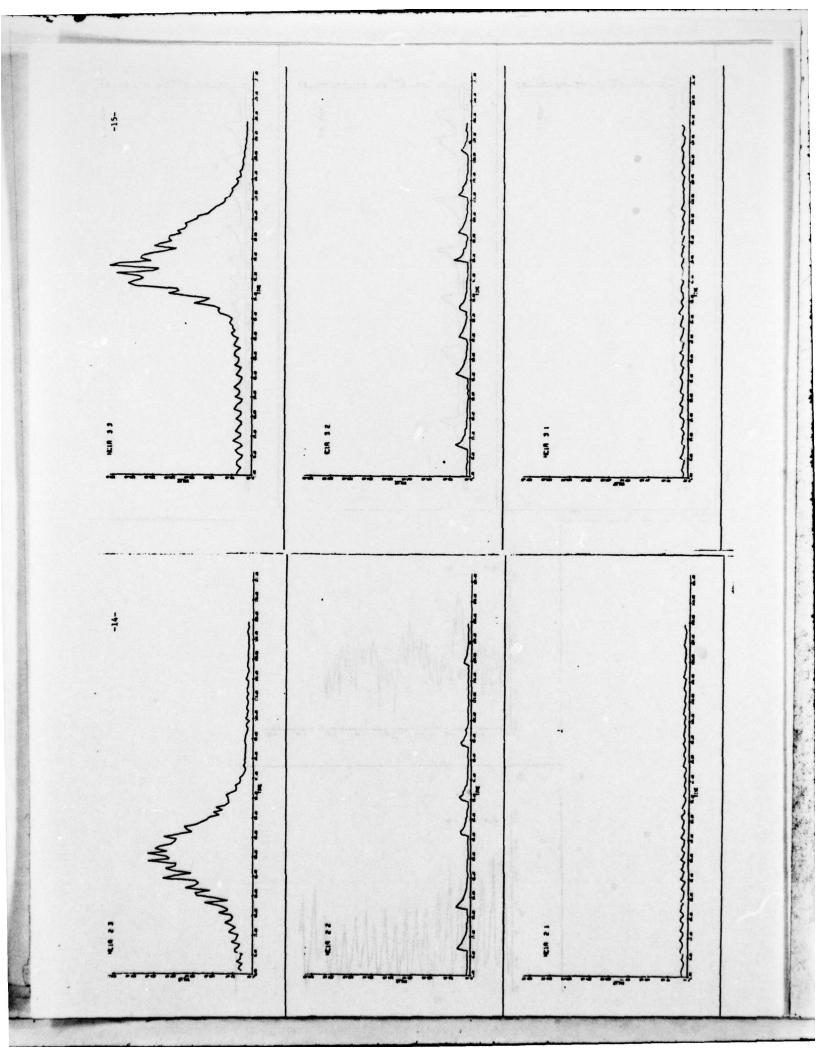


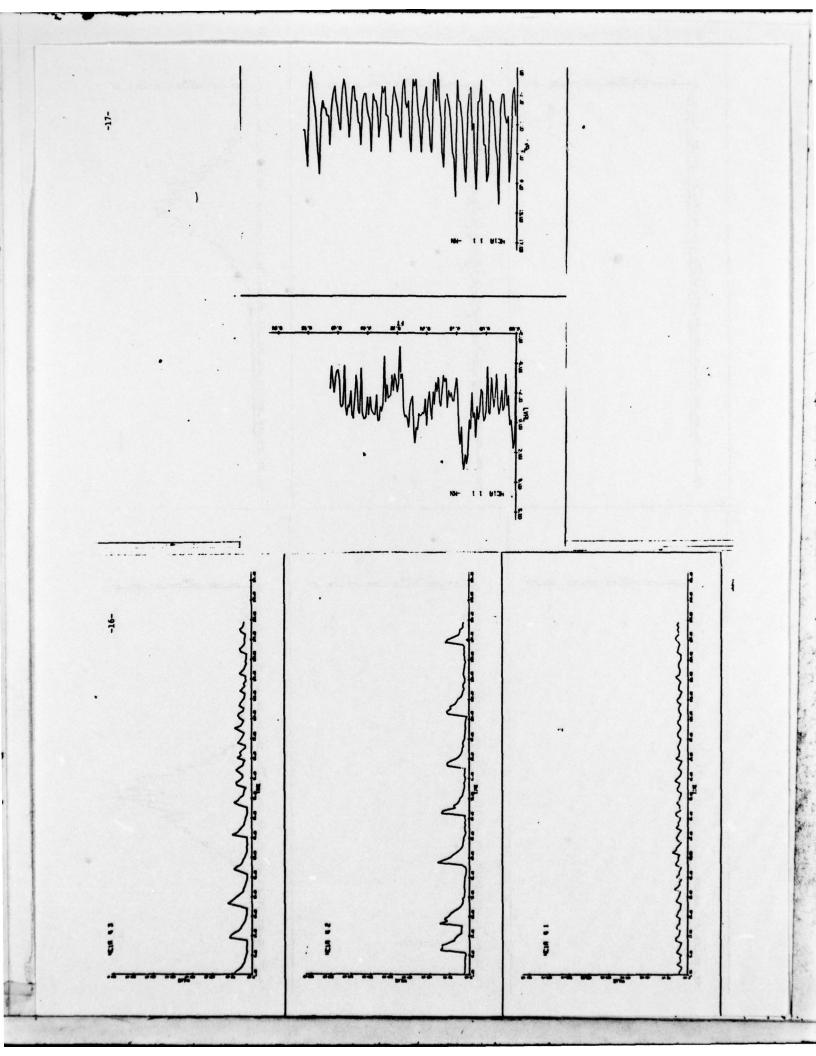


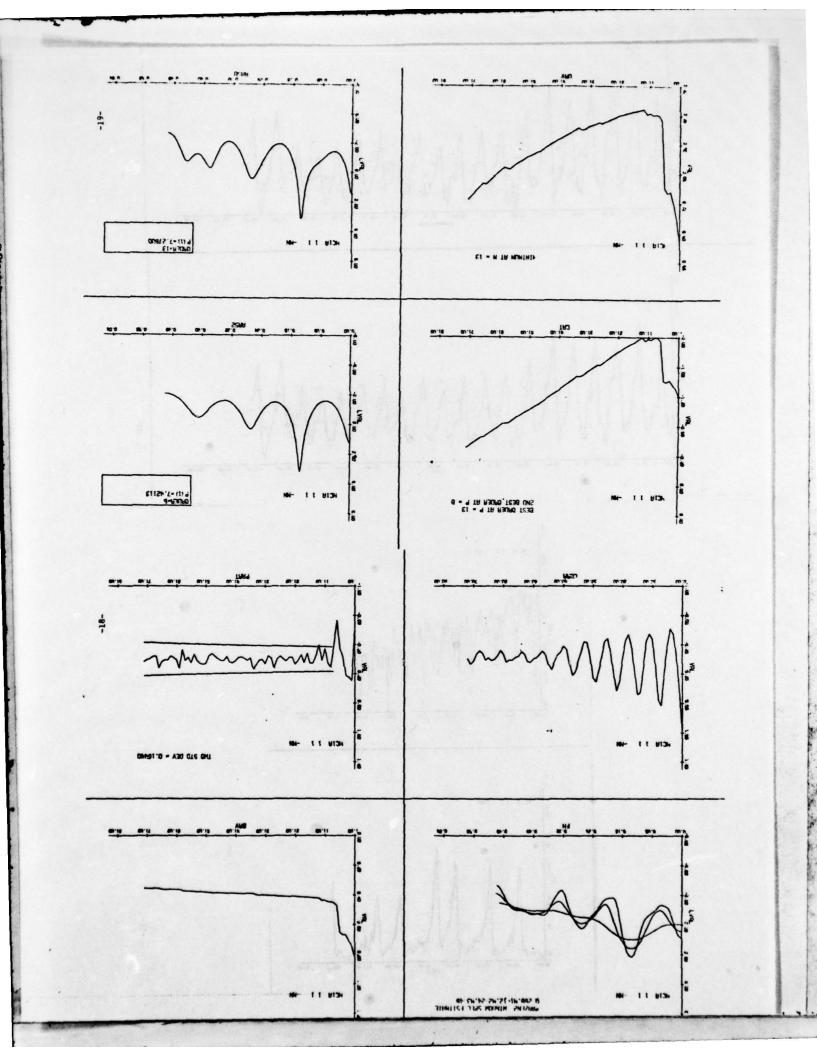


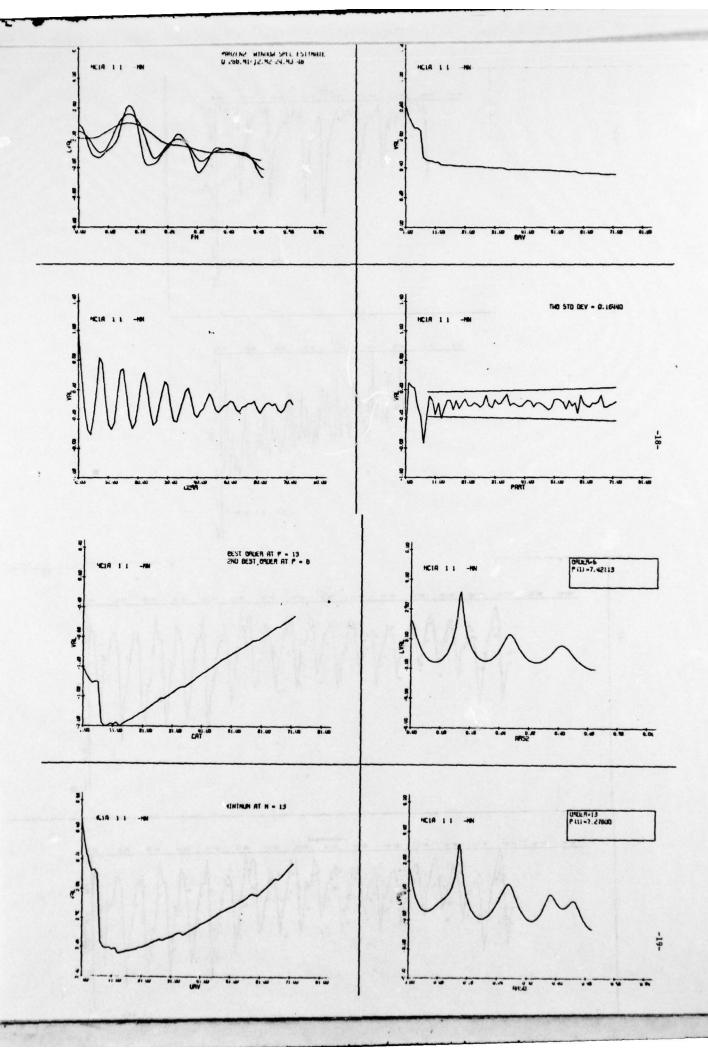


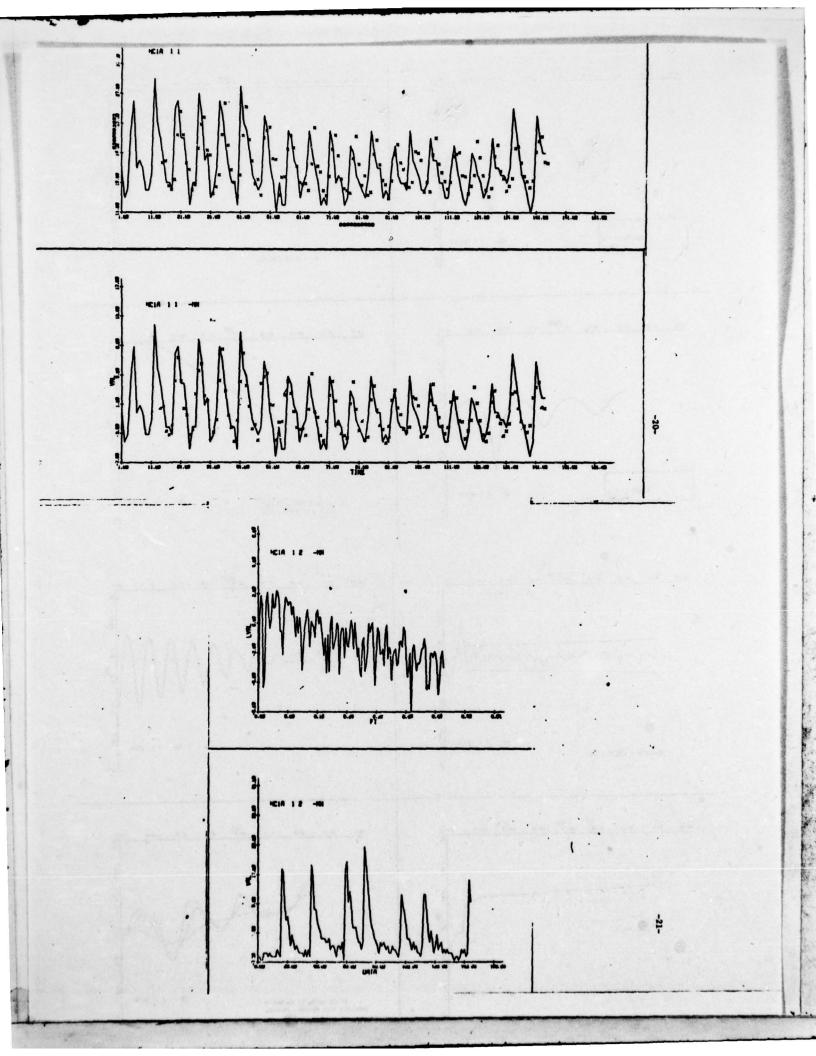


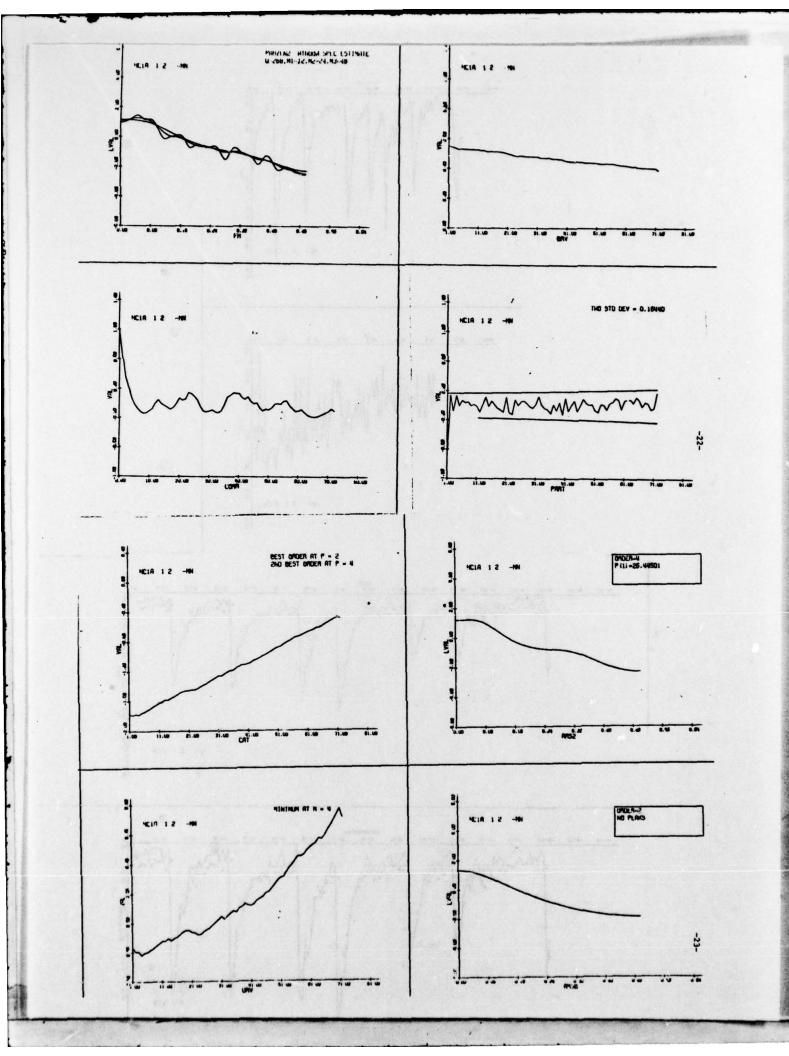


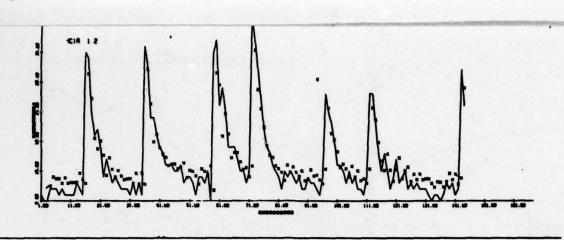


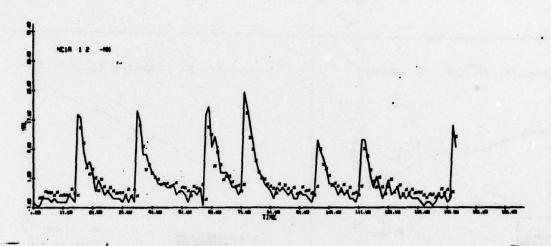




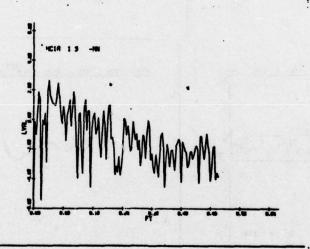


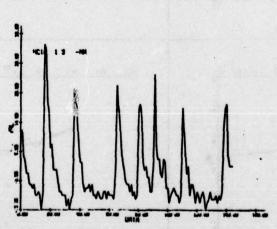


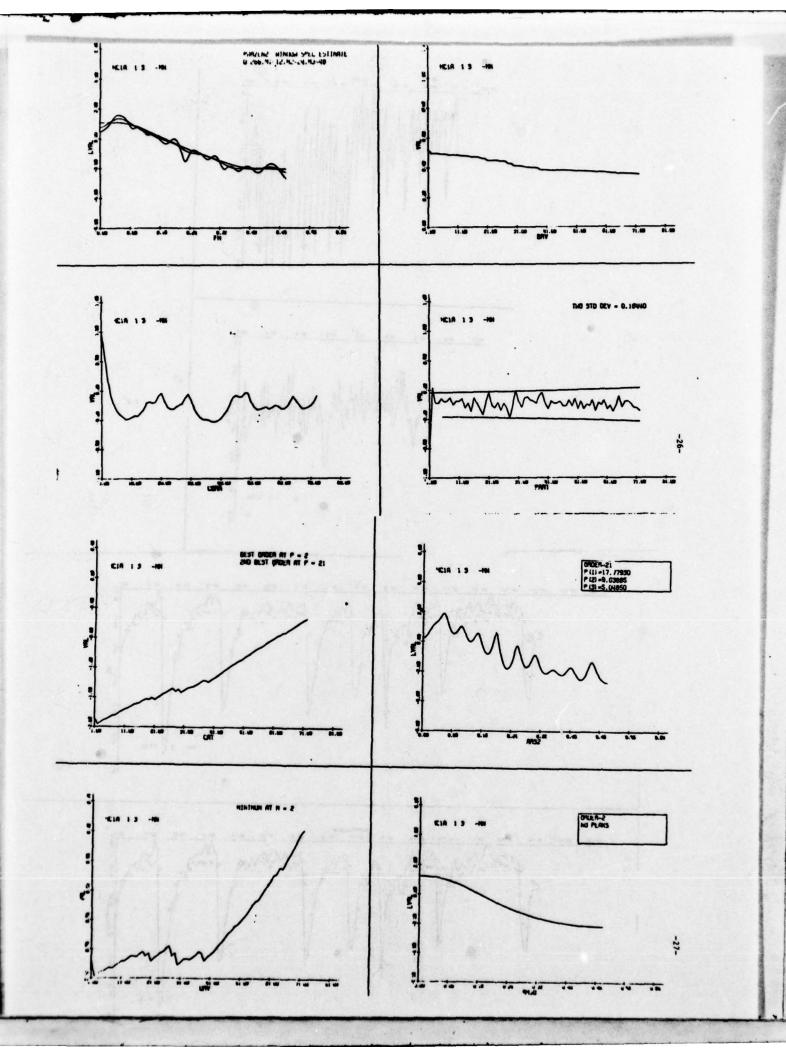


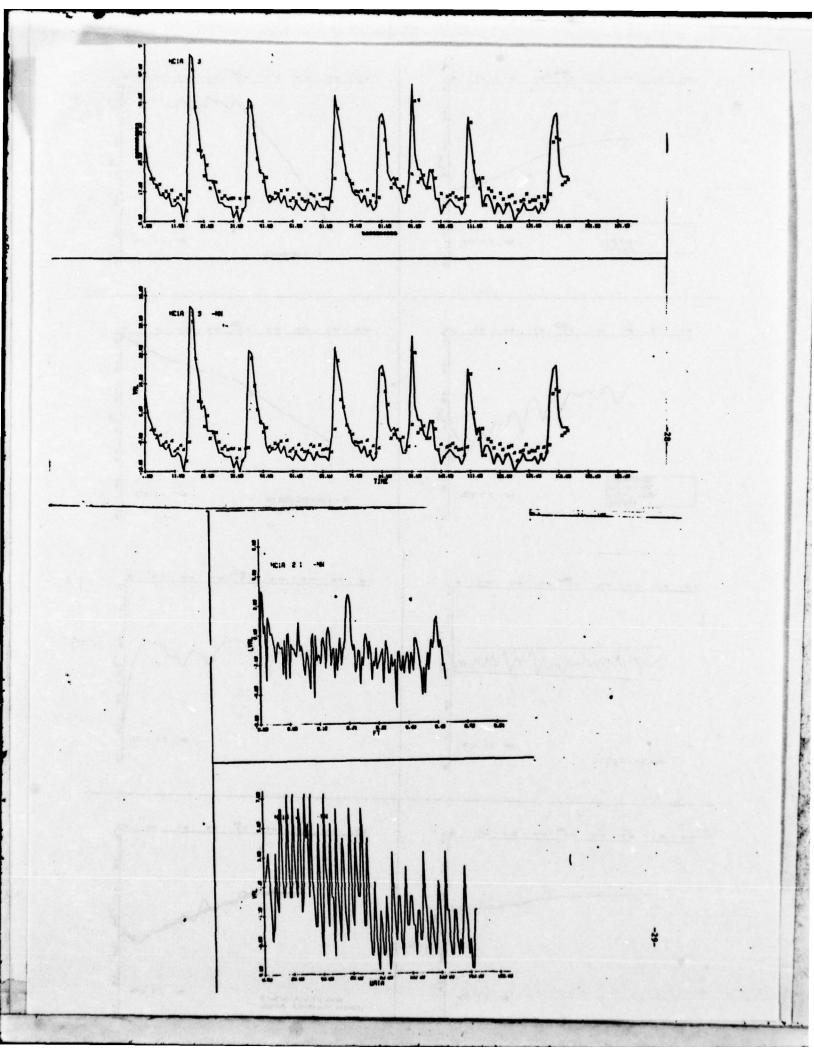


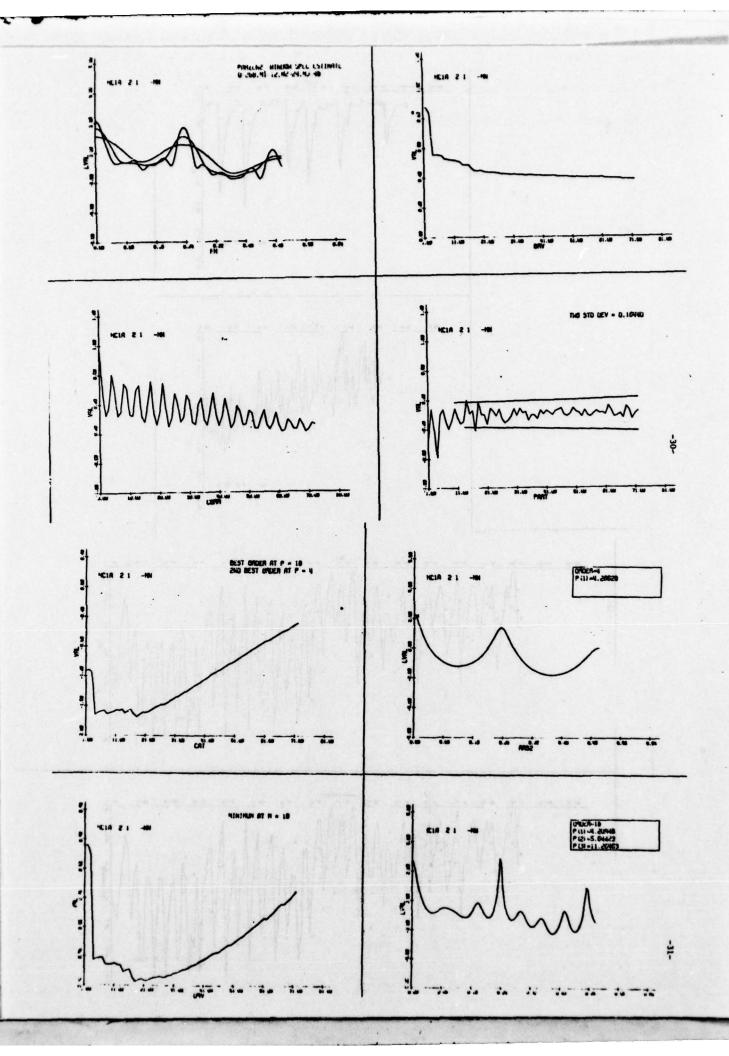
-24-

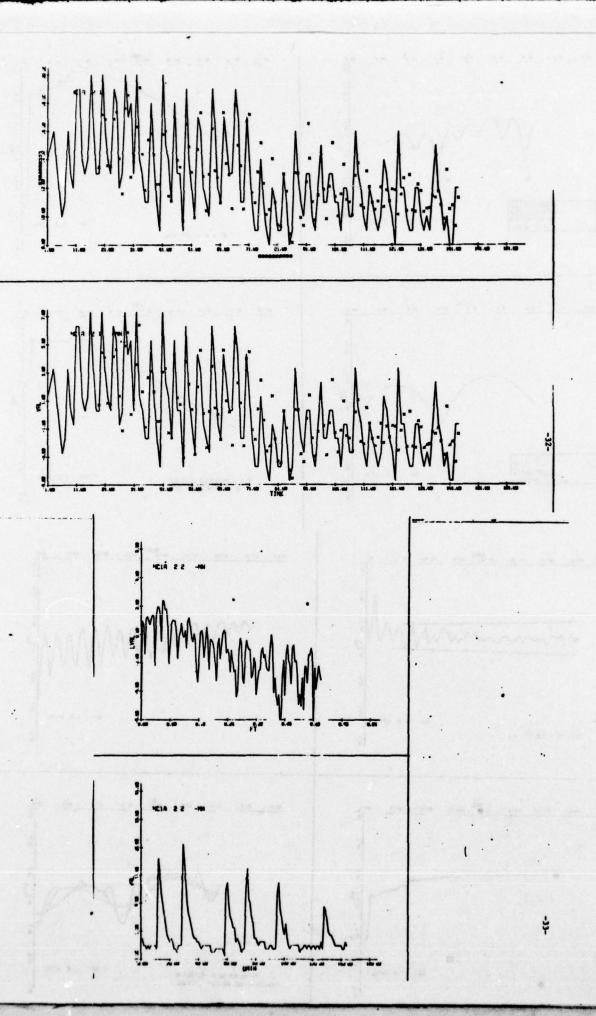


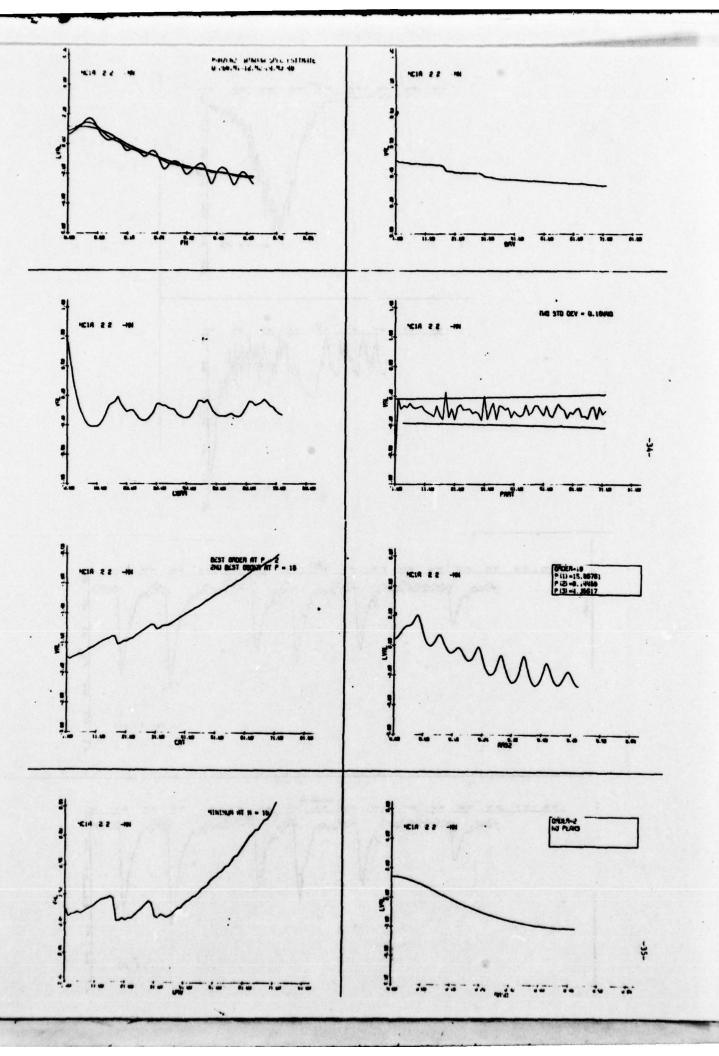


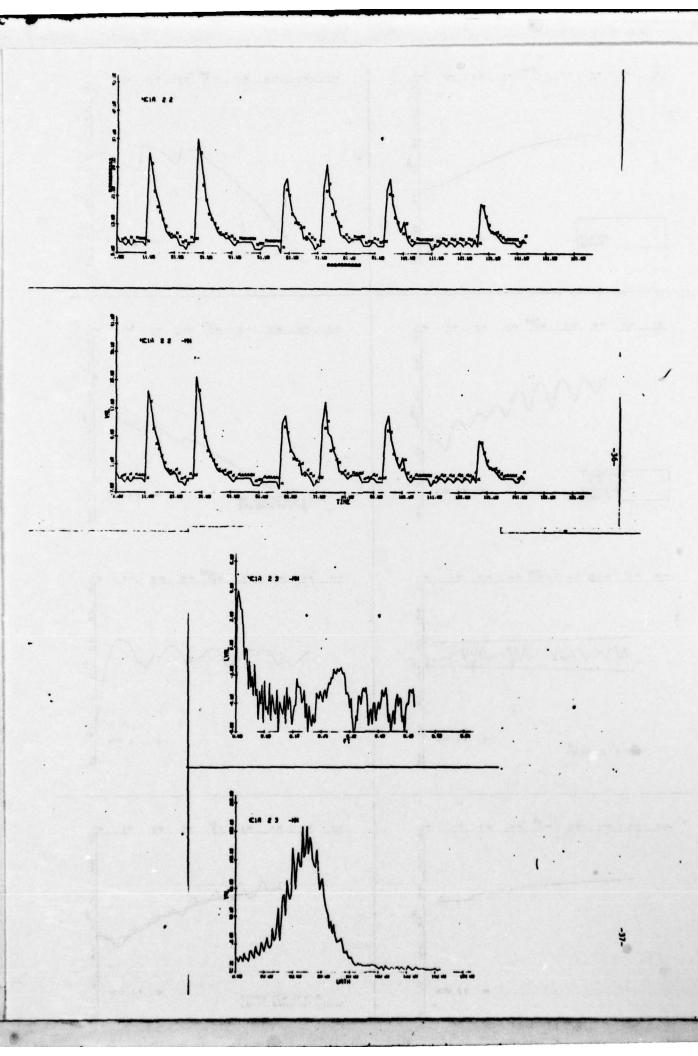


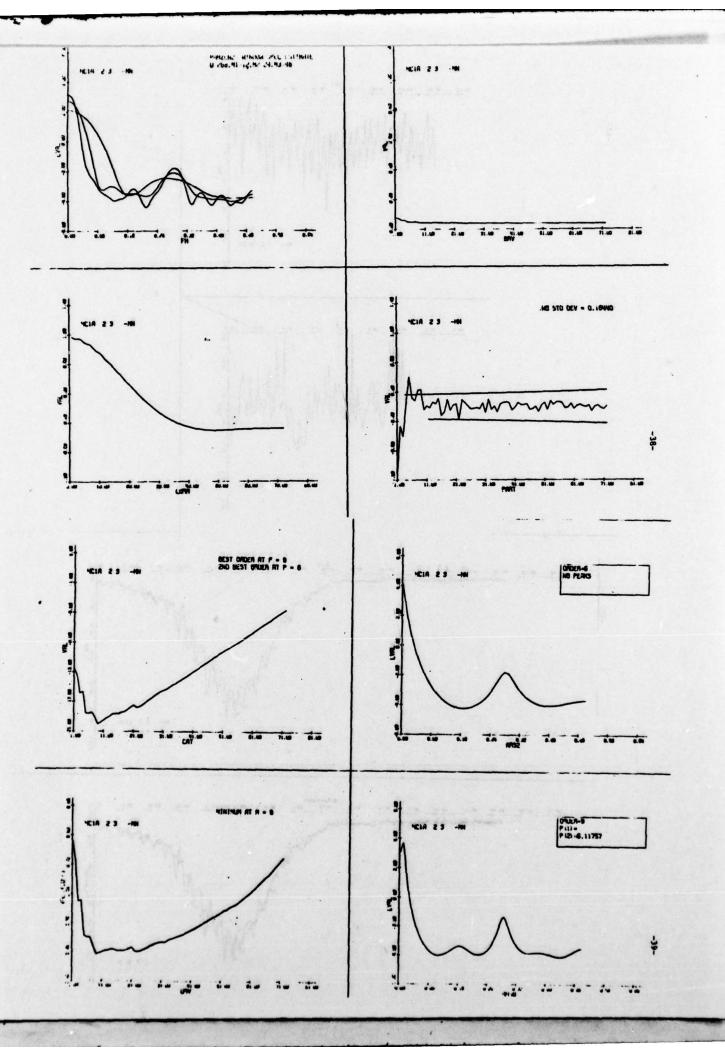


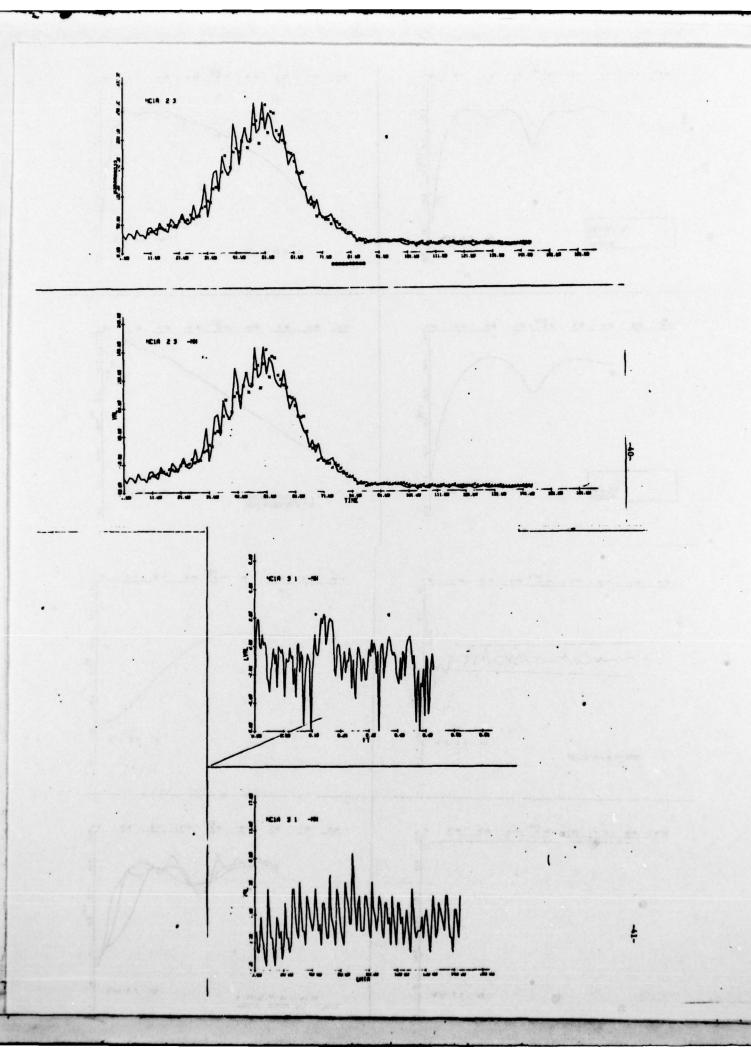


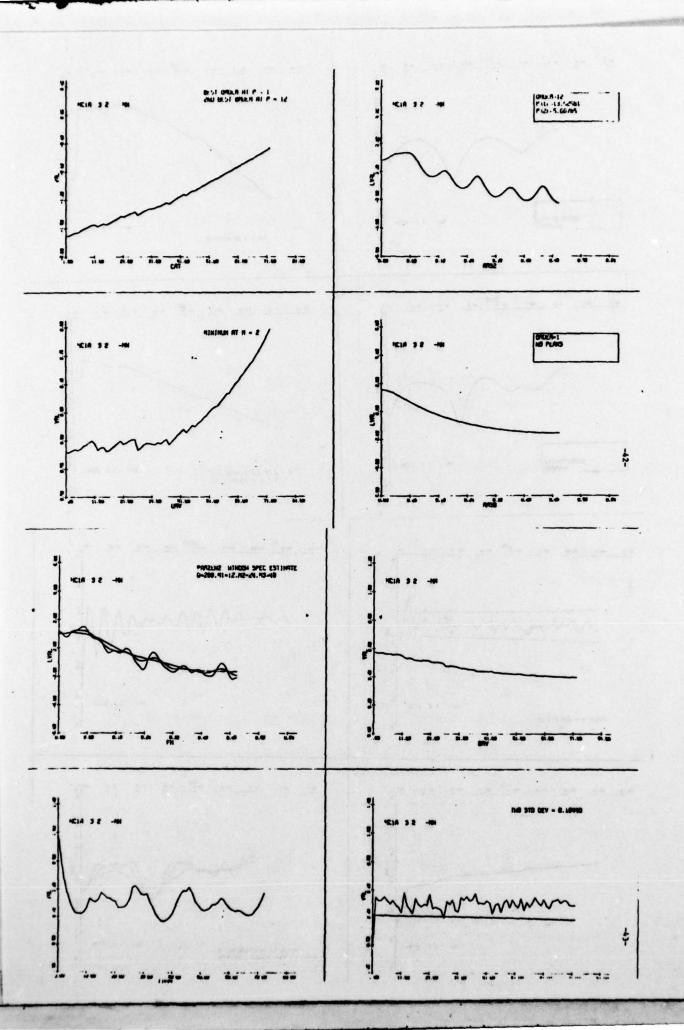


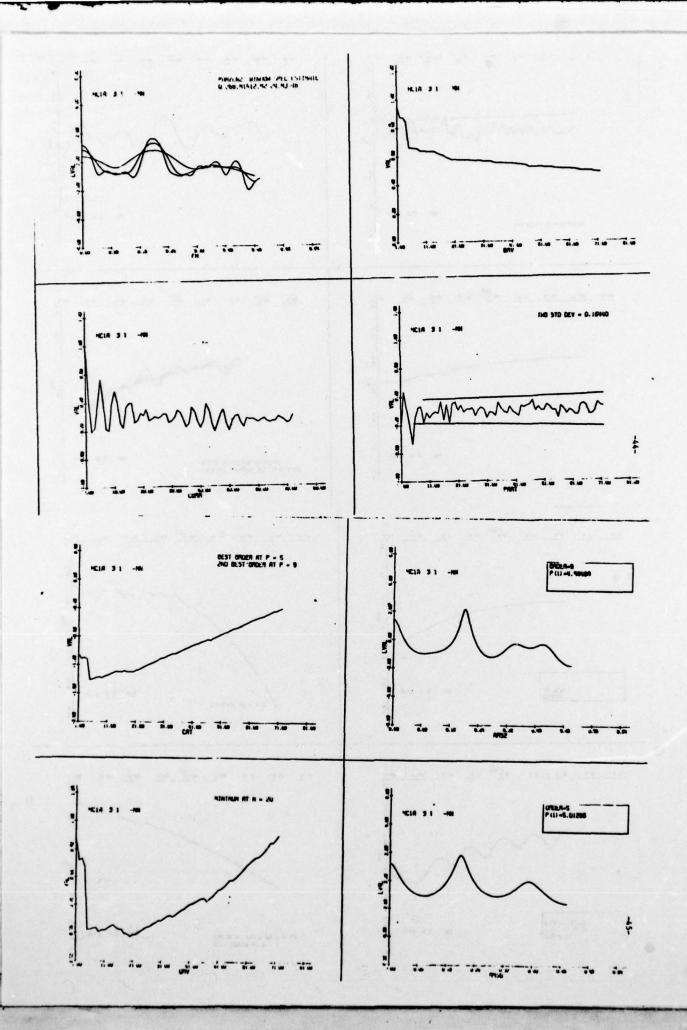


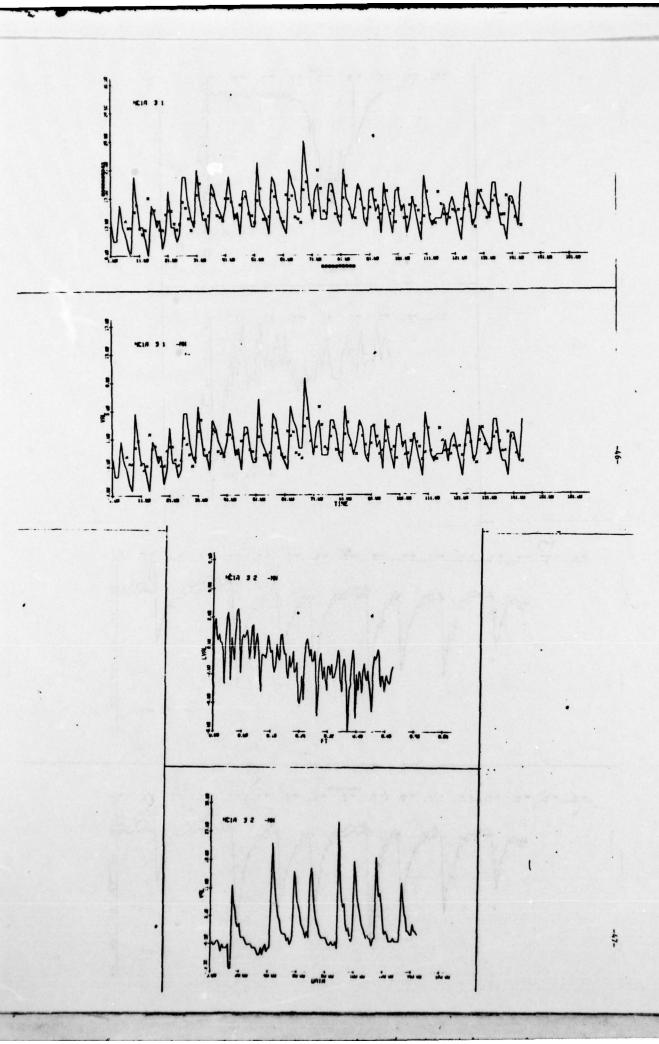


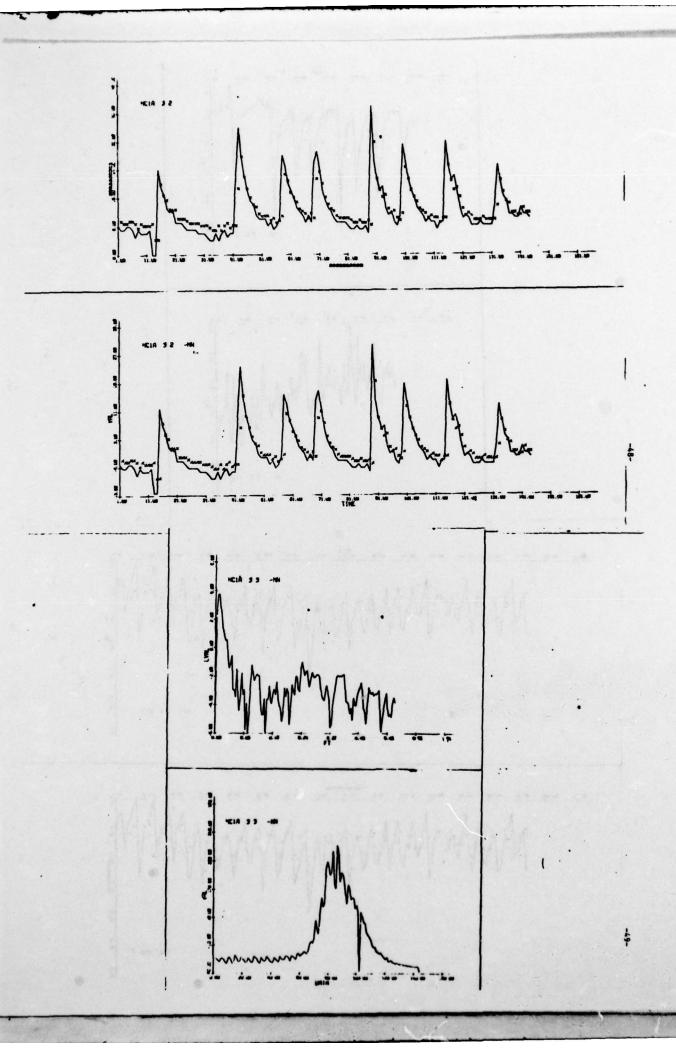


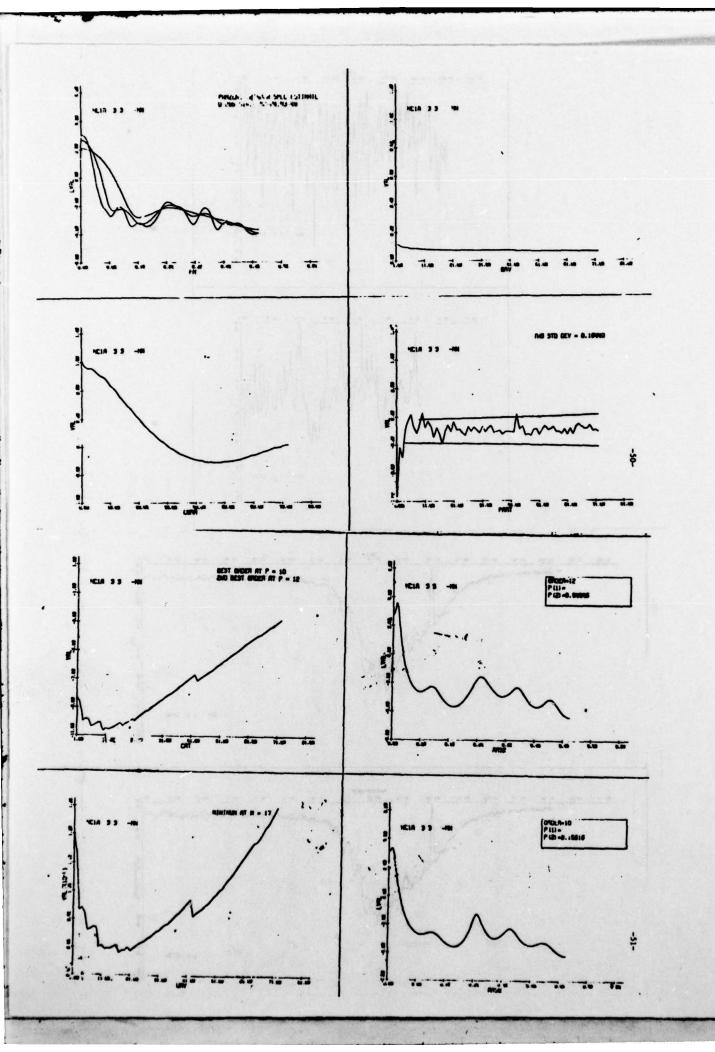


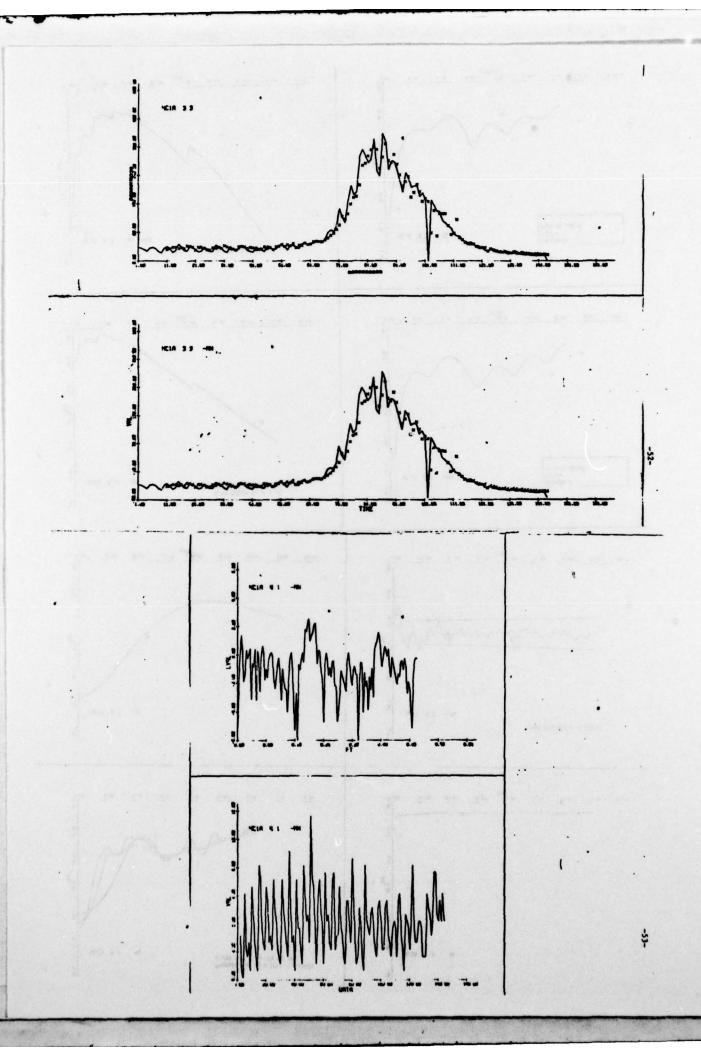


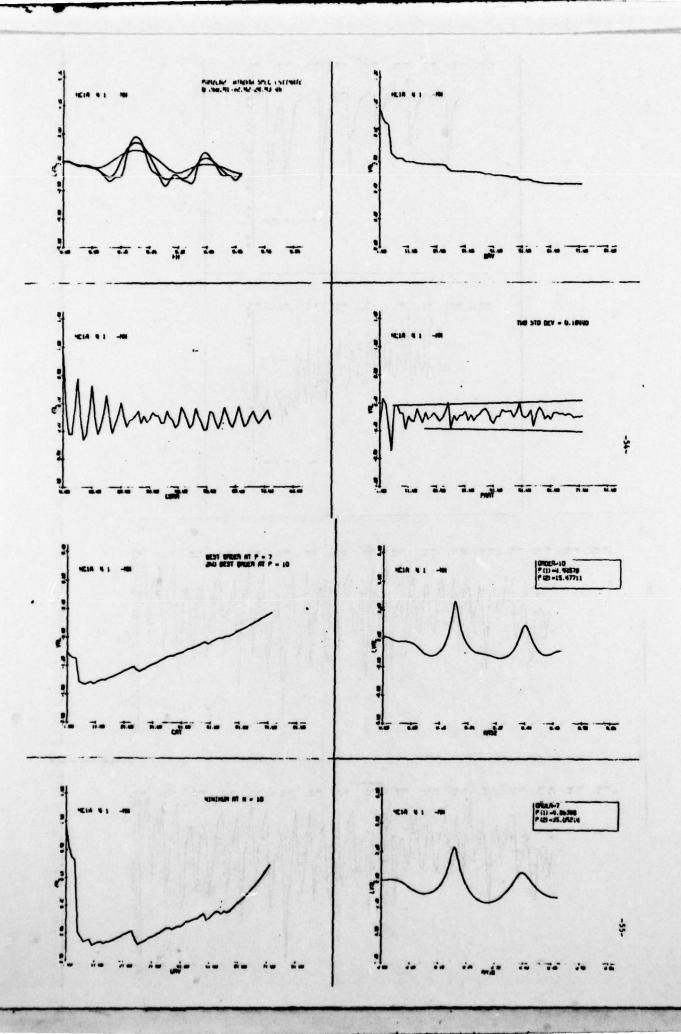


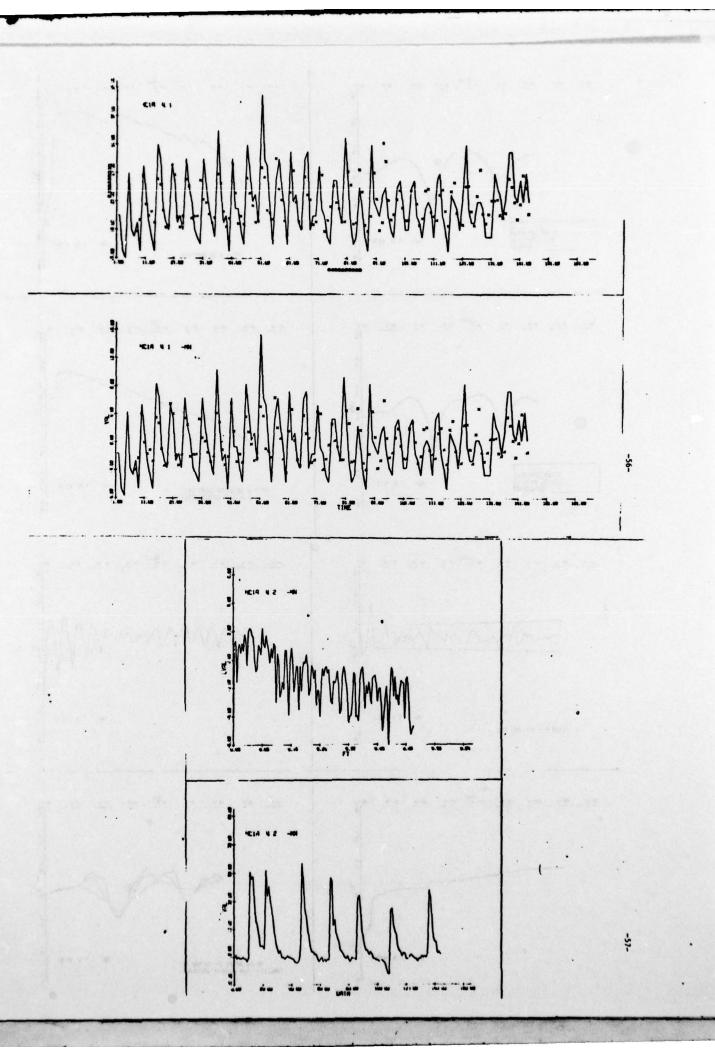


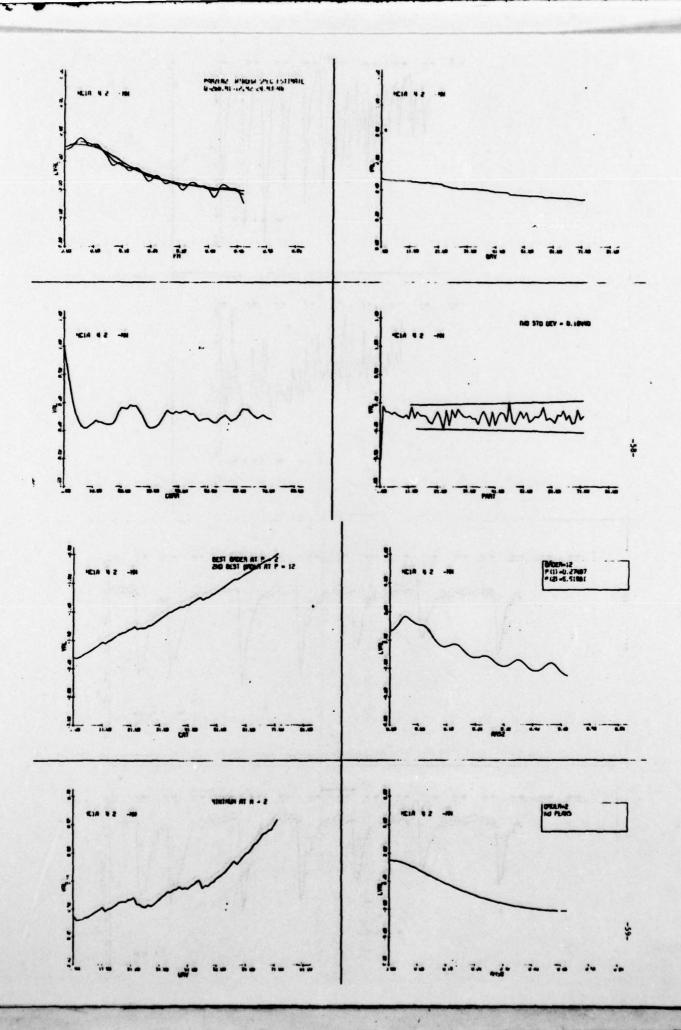


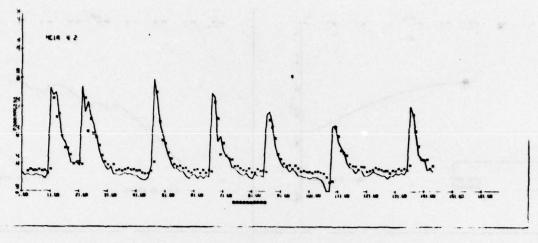


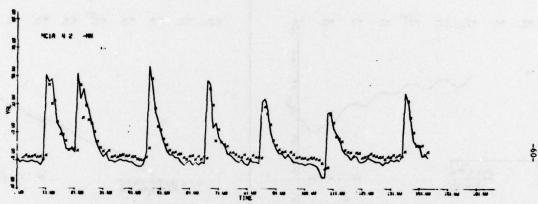


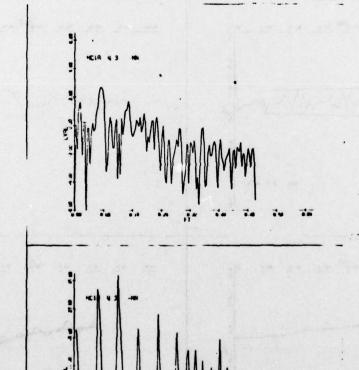




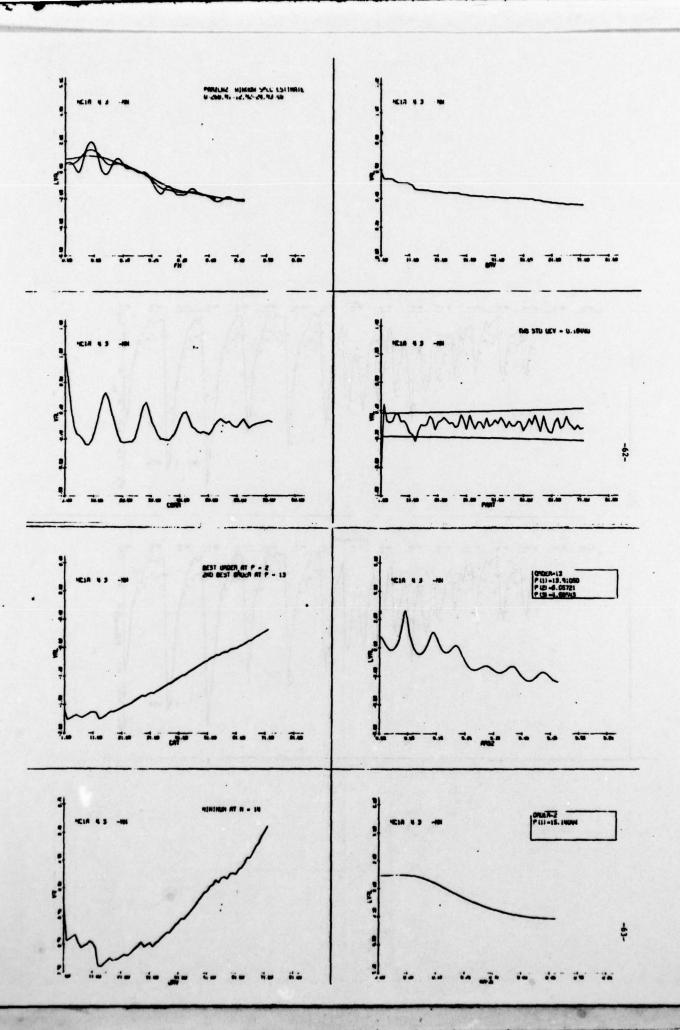


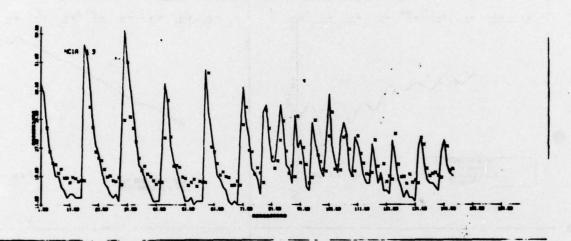






-61-





.

